(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 May 2003 (01.05.2003)

PCT

(10) International Publication Number
WO 03/035895 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/US02/34570

(22) International Filing Date: 28 October 2002 (28.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/330,669 26 October 2001 (26.10.2001) US
60/369,945 4 April 2002 (04.04.2002) US

(71) Applicant *(for all designated States except US)*: IMMUNIVEST CORPORATION [US/US]; 1105 North Market Street, Suite 1300, P.O. Box 8985, Wilmington, DE 19899 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: O'HARA, Shawn, Mark [US/US]; 1519 Isaacs Court, Ambler, PA 19002 (US). ZWEITZIG, Daniel [US/US]; 675 E. Street Road, Apt. 1407, Warminster, PA 18974 (US). FOULK, Brad [US/US]; 306 Brooke Road, Royersford, PA 19468 (US).

(74) Agents: ACETO, Joseph et al.; Immunicon Corporation, 3401 Masons Mill Road, Suite 100, Huntingdon Valley, PA 19006 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/035895 A2

(54) Title: MULTIPARAMETER ANALYSIS OF COMPREHENSIVE NUCLEIC ACIDS AND MORPHOLOGICAL FEATURES ON THE SAME SAMPLE

(57) Abstract: A highly sensitive assay is disclosed which utilizes a method for gene specific primed amplification of mRNA libraries from rare cells and rare transcripts found in blood.. The assay allows detection of rare, mRNA (10 copies/cell) found in 1 to 10 cells isolated through immunomagnetic enrichment. The assay is an improvement over multiplex PCR and allows efficient detection of rare coding sequences for circulating carcinoma cells in the blood. The methods are useful in profiling of cells isolated from tissues or body fluids and serves as an adjunct to clinical diagnosis of diverse carcinomas including early stage detection and classification of circulating tumor cells. Monitoring of nucleic acid and protein profiles of cells either in conventional or microarray formats, facilitates management of therapeutic intervention including staging, monitoring response to therapy, confirmation of remission and detection of regression.

Title: Multiparameter analysis of comprehensive nucleic acids and morphological features on the same sample.

This application claims priority under 35 U.S.C. §119(e) to U.S. Provisional Applications, No: 60/369,945 (filed April 4, 2002) and 60/330,669 (filed October 26, 2002), and PCT/US02/26867 (filed August 23, 2002).

Background of the Invention

Field of the Invention

This invention relates generally to gene specific amplification, analysis and profiling of cytosolic biomolecules useful in the fields of oncology and diagnostic testing. The invention is particularly useful in such fields as cancer screening, selecting and monitoring for chemotherapy treatment, or cancer recurrence. More specifically, the present invention provides methods, apparatus, and kits to facilitate comprehensive analysis of mRNA and DNA from tumor cells, or other rare cells from biological samples while simultaneously maintaining cell integrity for enumeration and morphological image analysis. To accomplish this, the invention also provides methods that permit the analysis of soluble cytosolic biomolecules releasable from a cell, such as a tumor cell, by means of permeabilizing reagents for determining expression profiles of the released nucleic acids, while still maintaining the morphological and antigenic characteristics of cells for subsequent or parallel multiparameter flowcytometric, image, and immunocytochemical analyses (see US 6,365,362). The invention also provides methods that enable the same comprehensive analyses using stabilized samples from aldehyde and aldehyde-urea derivative based fixatives.

Description of Related Art

Any given cell will express only a fraction of the total number of genes present in its genome. A portion of the total number of genes that are expressed determine aspects of cell function such as development and differentiation, homeostasis, cell cycle regulation, aging, apoptosis, etc. Alterations in gene expression decide the course of normal cell development

and the appearance of disease states, such as cancer. The expression of specific genes will have a profound effect on the nature of any given cell. Accordingly, the methods of analyzing gene expression, as such as those provided by the present invention, are important in basic molecular biological research and in tumor biology. Identification of specific genes, especially rare genes, can provide a key to diagnosis, prognosis and treatment for a variety of diseases that reflect these expression levels (Levsky, et al., Single-Cell Gene Expression Profiling, Science, 297:836-840, (2002)).

Differential gene expression is a commonly used method of assessing gene expression in a cell. In particular, cDNA microarray analysis compares cDNA target sequence levels obtained from cells or organs from healthy and diseased individuals. These targets are then hybridized to a set of probe fragments immobilized on a membrane. Differences in the resultant hybridization pattern are then detected and related to differences in gene expression of the two sources (US 6,383,749). This procedure requires slow and time-consuming analysis of several hundred thousand gene-specific probes. In addition, competing events such as interactions between non-complementary target sequences nonspecific binding between target and probe and secondary structures in target sequences may interfere with hybridization resulting in a decline in the signal-to-noise.

Gene specific primer sets have been described in assaying differential expression (US 5,994,076 and US 6,352,829). Here, gene specific primer sets were used to specifically amplify mRNA library subsets in complex libraries achieving a cDNA array signal improvement when compared to whole library labeling amplification. The focus of this type of analysis was to compare sample array expression profiles as part of gene discovery research, not development of methods for practical cellular RNA analysis with utility in diagnostics.

Hence while gene specific primer sets have been used to selectively amplify a specific subset of mRNA from an mRNA library, there exists a clear need to reduce the signal-to-noise ratio in an amplification process which is especially applicable in rare cell detection for diagnostic therapy to encompass both quantitative and qualitative analysis.

It is now generally accepted that the presence of circulating tumor cells (CTC) in a patient's blood provides an early detection system in assessing the need for therapeutic intervention. Highly sensitive assays to allow accurate enumeration of circulating carcinoma cells have shown that the peripheral blood tumor cell load correlate with disease state (Terstappen et al., Peripheral Blood Tumor Cell Load Reflects the Clinical Activity of the Disease in Patients with Carcinoma of the Breast, International J. of Oncology., 17:573-578, 2000).

Additionally, classification of cell type and origin would provide a more comprehensive platform for treatment. Emerging treatment for several cancers such as Diffuse Large B Cell Lymphoma (DLBCL) is based upon two different disease types correlating to a clinical prognosis (Rosenwald, et al., Use of Molecular Profiling to Predict Survival After Chemotherapy after Diffuse Large-B Cell Lymphoma, New England Journal of Medicine, 346:1937-1947, (2002)). In DLBCL, tumors originating from the germinal center B-cells are sensitive to chemotherapy and have a much higher chance of survival, while those from activated B cells tend to be more difficult to treat. These cell subtypes are thus dependent on the origin of the tumor cell.

Stratification of these subtypes is dependent upon the tumor's cell of origin. While in a few cases differences in subtypes can be determined by analysis of a single gene, entire arrays of combinations of genes are more determinative. Charting gene expression patterns through microarray analysis of gene expression levels would be a desirable indicator of tumor properties in other diseases such as lymphomas, acute leukemia, breast cancer, lung cancer and liver cancer. However, to adapt this genetic information for diagnostic use requires resolution of inherent significant signal-to-noise issues in present state-of-the-art technology.

Thus, there is great interest in the development of new methods for analyzing gene expression, especially where such methods provide for fast hybridization, highly specific binding of targets to complementary probes, and substantially improved signal-to-noise ratios. In addition, these methods have additional importance when assessing gene expression as it relates to cancer

and disease related states (see US App. 10/079,939 and US App. 09/904,472 both of which are fully incorporated by reference herein).

Summary of the Invention

The present invention provided methods, apparatus, and kits for assessing gene expression in amplified mRNA isolated from circulating rare cells (see Figure 1) which overcome the disadvantages of the prior art techniques which are described above. The present invention provides methods of isolating soluble or releasable cytoplasmic biomolecules from a single target cell or a cell population while maintaining structural cell integrity or phenotypic characteristics. Accordingly, cell(s) are either fresh or stabilized and fixed with a cross-linking agent, contacted with a pore-forming permeabilization composition, and the nucleic acids recovered. RNA from stabilized cells (PCT/US02/26867) is recovered via combinations of proteinase and nucleophiles reversal agents either for amplification and subsequent qualitative and quantitative PCR analysis or for quantitative analysis via gene specific subsets of reverse transcription (RT) primers fused with and followed by universal primer PCR amplification. Thus, one focus of the present invention prepares the cells for both cytoplasmic biomolecule analysis and phenotypic cell analysis by stabilizing and fixing cells prior to permeabilization and then releasing nucleic acids from the same stabilized cells.

The present invention is also directed to separating nuclear and/or mitochondrial DNA, RNA, proteins and other soluble components within a targeted cell by contacting a cell or cell population with a permeabilization composition and separately analyzing the released and/or unreleased fraction for one or more constituents such as the nuclear and/or mitochondrial DNA, total RNA, mRNA, soluble proteins, and other target substances (US App. 60/330,669).

The present invention incorporates the analysis of both cytoplasmic biomolecules and membrane or surface biomolecules from the same cell(s) or cell population by contacting the cell(s) with a permeabilization composition and separately analyzing the cytoplasmic biomolecules and the surface biomolecules to generate functional cell profiles encompassing characteristics

derived from genotypic and phenotypic cell characteristics for differentiating normal from transformed cells.

The isolation and rare cell analysis of the present invention are combined to provide the methods and reagents enabling comprehensive profiling mRNA acquired from rare cells. For example, those populations of cells defining circulating tumor cells (CTC) would be a type of rare cell found in peripheral blood and bone marrow of cancer patients. The mRNA is obtained through the cell preparation described in the present invention, but could also incorporate any protocol commonly used in the art.

After isolation and purification of mRNA from a sample containing the cells of interest, detection of extremely rare cell events with low mRNA copy numbers is achieved through gene specific RT-PCR panels with or without T7 RNA polymerase (T7RNAP) based pre-amplification procedure (US App 60/369,945). Pre-amplification is completed by linear amplification of the entire mRNA library using modifications of the Eberwine aRNA method (Van Gelder et al. 1990). In a preferred embodiment, generation of an anti-sense mRNA library (aRNA) library preamplificaiton results in at least a thousand-fold increase of all messages present in the original mRNA isolated from ferrofluid enriched circulating cells. Gene specific primers are then used to amplify only the gene panel of interest. These primers are designed to amplify transcripts indicative of known rare events like circulating tumor cells. The number of target sequences can be as small as two or as large as necessary to allow correlation with some indicative characteristic of the rare event. This can occur as separate individual reactions or within a single reaction vial. Subsequent analysis yields at least a qualitative assessment of the target sequences and is achieved with methods such as, but not limited to, one of two types of multigene analysis methods we present here as gene specific primed (GSP) arrays and/or GSP sets-RT (universal PCR).

Universal PCR achieves multigene analysis from sample recovered mRNA in a single reaction tube with or without mRNA library preamplification. No preamplification allows only one panel of genes to be analyzed at one time. Preamplification adds the advantage of analyzing a single sample in up to 1000 different reactions, thus many different panels of genes can be

interrogated at different times. While it will be noted that other methods are available, analysis of universal PCR cocktail panels is accomplished by array or capillary gel electrophoresis (CGE). The system allows, therefore, for both a quantitative and qualitative determination of 1 to thousands of separate mRNA types simultaneously when measured in cDNA microarray format.

Thus, the present invention includes a combination of the above mentioned isolation and profiling analysis directed to protocols and kits comprising some or all necessary reagents including a permeabilization composition, RNA recovery after cross-linking, magnetic microspheres with oligo(dT) probes covalently bound to the surface, and other gene specific magnetic microsphere-bound probes for capture and analysis of comprehensive RNA analysis using a small or large microarray, capillary gel electrophoresis (CGE), HPLC, electrophoresis and other analytical platforms.

Brief Description of the Drawings

Figure 1 shows a flowchart depicting the variety of capabilities and options enabled by the inventions described in this application for multiparameter analysis on a single sample. Phenotypic and genotypic analysis is obtained on fixed or non-fixed cells.

Figure 2 shows the reverse image (negative) of denatured total RNA analyzed by 2% agarose gel electrophoresis after SYBR Gold staining of about 1600 SKBR3 breast cancer cells that were Immuniperm-treated for various times before total RNA was isolated from the resultant supernatant via Trizol plus pellet paint co-precipitate.

Figure 3 shows a 1% denaturing total RNA agarose gel stained with ethidium bromide comprising whole cells, Immuniperm (saponin)-permeabilized cells from the cell pellet fraction, and cells from the supernatant fraction of Immuniperm®-permeabilized SKBR3 breast cancer cells.

Figure 4 shows a phosphor image of a Northern blot of the gel shown in Figure 3 hybridized by a polynucleotide kinase treated ^{32}P-labeled oligo(dT)

(25 mer) probe. The radioactive signals correspond to all the poly(A)+ mRNA transcripts of the total RNA which was derived from whole cells and the two Immuniperm® treated cell fractions from the gel shown in Figure 3.

Figure 5 shows a Northern blot from Figure 4 stripped through conventional dissociation and removal of the labeled oligo(dT) probe and re-probed with a nuclear-specific precursor rRNA probe.

Figure 6 shows the Northern blot, from Figure 5, that was stripped and re-probed with mitochondrial-specific 12s rRNA probe.

Figure 7 shows a cDNA array dot blot hybridization pattern comparison when the corresponding mRNA, used to generate the gel images in Figure 3, is alpha-^{32}P-nucleotide labeled during first strand oligo(dT) primed cDNA synthesis. Labeled first strand cDNA was then used as the hybridization probe. Pattern comparison shows the same relative abundance of mRNA exists in all three RNA cell fractions.

Figure 8 shows the gel image of the relative cytosolic total RNA, both quantitatively and qualitatively, obtained after separately treating multiple aliquots containing about 770 PC-3 cells each with Immuniperm®.

Figure 9 shows the preservation, recovery and RNA integrity analysis of 90-100% of the total RNA library using CytoChex™ and other aldehyde based fixatives followed by enzyme digestion. In this experiment, mass normalized portions of 300,000 SKBR3 cell line cells which were first spiked into freshly drawn 7.5ml peripheral blood (EDTA Vacutainer tube) in both control lanes without (phosphate buffered saline, PBS) and with three different fixatives being Cyto-Chex™, Stabilcyte™ and Transfix™. After mixing these were allowed to incubate at room temperature (20-25°C) for 24 hours. After which the SKBR3 cells were enriched from the blood using VU-1D9 (EpCAM)-ferrofluid immunomagnetic selection. The enriched cells from each treatment were split into 3 equal aliquots and treated with proteinase K (lane labeled "Post") or without proteinase K digestion (lanes labeled "Pre" for immediate

RNA isolation, and, lane labeled "No" meaning the only difference to Post is that No proteinase K component was added and all other manipulations are equal to Post). The resultant normalized RNA isolations were separated with a 1% denaturing agarose, stained with SYBR Gold, alpha imager densitometry imaged and then Northern Blotted and finally oligo(dT) probed to show relative quality and quantity of respective total RNA and mRNA libraries recovered.

Figure 10A & 10B shows relative rate of Cyto-Chex™, Stabilcyte™, Transfix™, paraformaldehyde, formaldehyde, glutaraldehyde and glyoxal fixation over a 1, 2, and 4 hour time course. . In this experiment, the relative-rate time course of Cyto-Chex™, Stabilcyte™, Transfix™, paraformaldehyde, formaldehyde, glutaraldehyde and glyoxal fixation were evaluated at 1, 2, and 4 hours. Samples of 7.5 ml of blood were prepared form a single donor by the same method as described in Figure 9. The only difference is that they were selected and processed for RNA isolation at 1, 2, and 4 hour time end points. The resultant normalized RNA isolations were separated with a 1% denaturing agarose, stained with SYBR Gold, alpha imager densitometry imaged and then Northern Blotted and finally oligo(dT) probed to show relative quality and quantity of respective total RNA and mRNA libraries recovered.

Figure 10C shows relative rate of fixation of Cyto-Chex™ vs. paraformaldehyde over a 15, 30, and 45 min time course. In this experiment, the relative-rate time course of fixation Cyto-Chex vs. paraformaldehyde were evaluated at 15, 30, and 45 min. Samples of 7.5 ml of blood were prepared from a single donor by the same method as described in Figure 9. The only difference is that they were selected at 15, 30, and 45 min time end points. Phosphoimaging quantitation of the oligo(dT) probed blots showing relative mRNA library quality and quantity showed the relative rate of: Formaldehyde = Paraformaldehyde (~4 fold) > Transfix™ (~2 fold) > Stabilcyte™ (~1.5 fold) > Cyto-Chex™ (~1x) = glutaraldehyde and glyoxal too slow to rank. The resultant normalized RNA isolations were separated with a 1% denaturing agarose, stained with SYBR Gold, alpha imager densitometry imaged and

then Northern Blotted and finally oligo(dT) probed to show relative quality and quantity of respective total RNA and mRNA libraries recovered.

Figure 11 shows the effects of variations on nucleophile and enzyme on the quality and quantity of RNA recovery from Cyto-Chex™ preservation supporting that when used in combined treatments for improved sequence analysis quality is likely. The resultant normalized RNA isolations were separated with a 1% denaturing agarose, stained with SYBR Gold, alpha imager densitometry imaged and then Northern Blotted and finally dT probed to show relative quality and quantity of respective total RNA and mRNA libraries recovered.

Figure 12A shows the feasibility of diagnostic applications demonstrated by detection of specific mRNA From 10 SKBR3 cells/7.5ml Blood in 24hr Cyto-Chex™ stabilized blood with proteinase recovery and aRNA preamplification. Feasibility of diagnostic applications are here demonstrated by sensitive and reproducible detection of specific mRNA from triplicate 10 or 20 SKBR3 cells spiked into 7.5ml peripheral blood. The spiked blood was stabilized immediately treated with Cyto-Chex to stabilize the cellular RNA. After incubating for one day at room temperature (20-25C) the stabilized cells were selectively enriched using VU-1D9 (EpCam)-Ferro Fluid Immunomagnetic selection. Enrichment was followed by proteinase K digestion to liberate the RNA so that silica binding RNA isolation followed by aRNA preamplification and gene specific quantitative RT-PCR could be performed for CK19 and EpCAM.

Figure 12B and 12C shows CK19 and EpCAM respective Q-PCR from SKBR Cell Spike, ferrofluid selection, and CytoChex™ treatment, followed by proteinase reversal. This experiment shows the results of the quantitative RT-PCR analysis, which was normalized to original total RNA mass prior to graphing. Thus, the values shown are equivalent to the original mRNA population contained in the original RNA isolation prior to aRNA amplification.

Figure 12D and 12E shows CD19 and EpCAM respective Q-PCR on aRNA derived from SKBR cells treated with cell stability reagents. These experiments show the RNA derived from the three different fixatives, which were shown in Figure 9 Cyto-Chex™, Stabilcyte™ and Transfix™. These proteinase K recovered RNA samples were also here subjected to the same mRNA template analysis of aRNA preamplification and gene specific quantitative RT-PCR for CK19 and EpCAM as shown in Figures 12B and 12C respectively. These data shown here are normalized to both total RNA mass, which is equal to the aRNA mass yield in this experiment. Thus this fixative derived relative RNA quantity and quality comparison is a measure of both aRNA (Figure 12A) and followed by the quantitative RT-PCR shown here. Both of these comparisons are a measure of the respective fixative dependent RT template quality after the RNA was recovered using only the proteinase K method.

Figure 13A shows RT-PCR amplification efficiency of a CK19-cRNA standard containing the 3'-most 800 base sequence of the CK19 mRNA transcript. Serial two-fold dilutions of the CK19-cRNA standard containing 200, 100, 50, 25, 12,5 copies were spiked into 2 ng of CK19 negative total RNA from white blood cells in triplicate resulting in a maximum coefficient of variation of 27%. Standard deviation bars are shown. Dilutions of cRNA to less than one copy and no template controls did not produce detectable signals.

Figure 13B shows the relative RT-PCR gene expression levels after agarose gel electrophoresis. CK19 cRNA was spiked into total RNA from white blood cells at levels of 25 copies, 250 copies, 2,500 copies and 25,000 copies in panel 1, panel 2, panel 3, and panel 4, respectively.

Figure 13C compares the relative representation in the same mRNA library of unamplified and T7 promoter-based amplified mRNA transcripts. Relative abundance was assessed by examining 8 different mRNA transcripts (PSA, PSM, MGB1, MGB2, CK8, CK19, PIP, EpCam) using the RT-PCR kinetic curve method.

Figure 14A shows scatter plot bar graphs of a survey of genes indicating the presence of circulating epithelial cells. Human blood, immunomagnetically enriched for cells expressing the EpCAM antigen on their cell surface, the samples were first aRNA preamplified and then 25ng were reverse transcribed. Aliquots were then analyzed by agarose gel electrophoresis after RT-PCR on a select group of genes. 13 healthy donors (7 male, 6 female) are designated as the N column for each gene measured and 9 serially sampled HRPC patients containing circulating tumor cells (CTC) were determined by flowcytometry and are designated as the P column for each gene measured. Horizontal lines in each column indicate threshold values above which true positives were counted.

Figure 14B depicts a survey of genes indicating prostate tumor organ of origin status via the same methods as described in 14A.

Figure 14C depicts a survey of genes indicating the presence of therapeutic target status via the same methods as described in 14A.

Figure 15A, 15B, and 15C show individual HRPC patient longitudinal monitoring of CTC and RT-PCR multigene analysis before, during, and after new line chemotherapy. The x-axis shows sampling time in weeks, the left y-axis shows the CTC level with the solid circle symbol. The right y-axis shows the relative mRNA expression levels with corresponding symbols of open-square for Androgen Receptor (AR), open circle for Hepsin (HPN) and open triangle for multidrug resistance (MDR1). Relative mRNA levels are illustrated here during treatment courses of Lupron alone as shown Figure 15A, and 2 patients being treated with Lupron combined with administration of doses of Taxotere and Estramustine chemotherapy symbolized by the vertical arrows on the x-axis (Tx/Ex) in Figure 15B and Figure 15C. Bars on top indicate long term hormonal ablation treatment was on going.

Detailed Description of the Preferred Embodiments

As has been indicated in the foregoing discussion, a more comprehensive and practical form of cancer diagnosis must also include analysis of intra- and extra-cellular membrane antigens as well as analysis of cellular RNA content and DNA content in the same cell or cell population, which up to now have been mutually exclusive processes (US 6,365,362). This exclusivity was due to the basic incompatibility of pre-analytical cell preparation methods for analyzing structural intracellular antigens, having the major objective to maintain cell integrity, with methods of isolating cytoplasmic biomolecules. Alternatively, pre-analytical cell preparation could also be limited to soluble cytoplasmic RNA, total cellular RNA, total cellular DNA, and/or proteins, having the major objective to homogenize cells in order to release soluble intracellular components (US 6,329,179). In particular, traditional phenotypic characterizations required fixation of cell structures achieved through exposure of cells to a cross-linking agent, such as paraformaldehyde, formaldehyde, glutaraldehyde, etc. These harsh cell fixation conditions simultaneously cause undesirable covalent crosslinking and/or fragmentation of all the isolatable RNA species. Similar intracellular DNA-protein cross-links have recently been reported (Quievryn and Zhitkovick, Loss of DNA-Protein Crosslinks from Formaldehyde-Exposed Cells Occurs Through Spontaneous Hydrolysis and an Active Repair Process Linked to Proteosome Function, Carcinogenesis, 21:1573-1580 (2000). So-called non-formaldehyde or non-paraformaldehyde fixatives (e.g., Cyto-Chex™Streck Labs, Omaha, NE) are cell-stabilizing additives containing formaldehyde-urea derivative donor compounds. It is used as a preservative for circulating tumor cells in blood during shipment or storage as disclosed in a co-pending application (PCT/US02/26867 incorporated by reference herein). However, the studies conducted by the present inventors have shown that even Cyto-Chex™, which contains only trace levels of free formaldehyde, apparently slowly releases formaldehyde that can cross-link intracellular RNA to intracellular proteins. Such cross-links were fully reversed by the methods of this invention to allow comprehensive RNA analysis. Cellular RNA and DNA analysis are therefore conventionally prepared on unfixed fresh cells or cells that are

preserved with reagents that do not cross-link or of which the cross-linking can be reversed during the mRNA release from the cells. RNAlater™ (Ambion) is commercially available RNA stabilization solution, which stabilizes RNA but does not allow immunomagnetic, immunochemistry or image analysis on the same sample and is not effective for blood. PreAnalytiX offers a blood RNA stabilizer but is nothing more than the chaotropic agent guanidine isothiocyanate solution (GITC) solution in a Vacutainer™ tube enabling nothing more than traditional homogenization based solely on total RNA isolation.

In general, mRNA recovered from fixed cells is not quantitative and is severely degraded or fragmented reducing the size of intact RNA with an average size of approximately 1750 bases as much as ten-fold to a highly variable average size of approximately 200 bases, and contains many complex chemical modifications, which are not well understood. However, the net effect of fixative derived RNA is severely compromised mRNA analysis (Current Protocols in Molecular Biology, Wiley, (2002)). Tedious non-quantitative mRNA salvage techniques combined with reverse transcriptase polymerase chain reaction (RT-PCR) analysis designed for amplicons of less than 100 base pairs in length show limited value, albeit in a qualitative not quantitative manner (US 5,346,994). Further, this limited RNA analysis of fixed cells must follow phenotypic analysis. Thus, the two processes cannot be run sequentially on the same cell sample, because traditional RNA isolation techniques require complete cell lysis or homogenization, destroying cell structure and further complicating analysis by intermingling the cellular DNA and RNA populations (Maniatis et al., Molecular Cloning-A Laboratory Manual, 2nd ed., Cold Spring Harbor Press (1989)). Previous reports have shown a need for improving methods of RNA recovery in tissue (Godfrey, et al., Quantitative mRNA Expression Analysis from Formalin-Fixed, Paraffin-Embedded Tissues Using 5' Nuclease Quantitative Reverse Transcription-Polymerase Chain Reaction, J. of Molecular Diagnostics, 2:84-91 (2000)). Applications of formaldehyde and urea based fixatives that stabilize recoverable quantitative, high quality full-length intact total and mRNA

libraries from blood and thus enabling comprehensive analysis are the basis of one aspect of this invention.

Quite unexpectedly, saponin, used as a permeabilizing agent, was found to be a highly selective and efficient releasing agent for intracellular cytoplasmic RNA and other biomolecules, thereby obviating the need for cell lysis or homogenization. This novel use of saponin as the RNA releasing agent of choice is a particularly advantageous component of the present invention. Surfactants such as saponin have traditionally been used to examine the expression of intracellular antigens by permeabilization of the cell membrane allowing for incorporation of staining reagents while maintaining cell integrity. For instance, analysis of chromosomes or genes by fluorescence in situ hybridization (FISH), or in staining of intracellular constituents, such as DNA in nuclei, with the nuclear stain, DAPI, or in immunostaining of cytokeratins with specific labeled antibodies play a critical role. Release of cytoplasmic intracellular proteins RNA or DNA generally is done by solubilization or complete lysis of the cells with stronger surfactants, such as Triton X-100. Saponin, however, has heretofore not been used to study both expression of soluble intracellular antigens including RNA and phenotyping of individual cells or cell populations in the same specimen. Accordingly, methods allowing sequential phenotypic analysis as well as analysis of intact RNA and soluble proteins in the cytoplasm of the same cell specimen are highly desired and are the subject of this invention.

Accordingly, the present invention provides advantageous methods, apparatus, and kits for the rapid and efficient RNA profiling of all cells and especially targeted cells found in biological samples. The present invention provides methods for allowing separate analysis of both phenotype and genotype. Phenotype is interrogated and profiled via antibody antigen protein and mass spectrometry profiling methods and comprehensive analysis of intact cytoplasmic RNA from the same cell or cell population. Genotyping of the sample genomic and mitochondrial DNA can be separately profiled by any means available to those skilled in the art. Similar to the amplification of the mRNA library, the respective genomic and mitochondrial libraries can be preamplified enabling numerous assays to be performed without loss of clinical sensitivity due to Multiple Displacement Amplification (MDA)

technology enables the first effective whole genome amplification method. MDA is a rapid, reliable method of generating unlimited DNA from a few cells.

The invention described herein may be used effectively to isolate and characterize cell phenotype, such as cell surface antigens, intra-cytoplasmic antigens and any type of RNA, and genotype. Both phenotypic and genotypic analysis can be performed sequentially on the exact same sample. For example after cell surface analysis and RNA harvesting, the remaining intact nuclei and mitochondria can be analyzed downstream by all standard RNA (mt RNA, hRNA), DNA and protein based analysis techniques such as S1 nuclease, ribonuclease protection, RT-PCR, SAGE, DD-RT-PCR, microarray cDNA hybridization, ISH, FISH, SNP, all RNA and all genomic-based PCR techniques and any protein analysis systems.

One of the many applications of this type of cell analysis is in cancer diagnostics. Many clinicians believe that cancer is an organ specific disease when confined to its early stages. The disease becomes systemic by the time it is first detected using methods currently available. Accordingly, evidence to suggest the presence of tumor cells in the circulation would provide a first line detection mechanism that could either replace, or function in conjunction with other tests such as mammography or measurements of prostate specific antigen. By analyzing cellular phenotype (protein and RNA) and genotype through specific markers for these cells, the organ origin of such cells may readily be determined, e.g., breast, prostate, colon, lung, ovarian or other non-hematopoietic cancers. Thus in situations where protein, RNA, and genome can be analyzed, especially where no clinical signs of a tumor are available, it will be possible to identify the presence of a specific tumor as well as the organ of origin. As these profiles define cell function, they also indicate what the most appropriate therapy type and course should be when used in cancer cell detection. Further in monitoring cases where there is no detectable evidence of circulating tumor cells as with post operative surgery or other successful therapies, it may be possible to determine from a further clinical study whether further treatment is necessary.

In order to provide for a more comprehensive and early diagnosis, one embodiment of the invention includes the methods for isolating cytoplasmic biomolecules from a cell or population of cells, contacting the cell or cells with

a permeabilization compound, and isolating the cytoplasmic biomolecule of interest from the cell while maintaining cell integrity for subsequent phenotypic and morphological analysis.

The targeted rare event in this invention refers to the expression of any biomaterial indicative, at least in part, to a known rare event. Accordingly, hormones, proteins, peptides, lectins, oligonucleotides, drugs, chemical substances, nucleic acid molecules (such as RNA and/or DNA) and bioparticles such as cells, apoptotic bodies, cell debris, nuclei, mitochondria, viruses, bacteria, and the like would be included in the embodiment of this invention.

The fluid sample includes, without limitation, cell-containing bodily fluids, peripheral blood, bone marrow, urine, saliva, sputum, semen, tissue homogenates, nipple aspirates, and any other source of rare cells that is obtainable from a human subject.

"Cytoplasmic biomolecules" includes cellular target molecules of interest such as, but not limited to, protein, polypeptides, glycoprotein, oligosaccharide, lipids, electrolytes, RNA, DNA and the like, that is located in the cytoplasmic compartment of a cell. Upon contacting a cell with a permeabilization compound and subsequent cell separation, the cytoplasmic biomolecules are present in the supernatant for downstream analysis. All soluble cytoplasmic biomolecules, for example, the entire cytoplasmic RNA library or target components capable of traversing the membrane pores can be isolated and analyzed. In a preferred embodiment, the focus is on the analysis of transcribed mRNA and translated proteins, for example in CTC, as indicators of oncogenic transformations of interest in the management of cancer diagnosis and therapy.

"Membrane biomolecules" includes any extracellular, intra-membrane, or intracellular domain molecule of interest that is associated with or imbedded in the cell membranes including, but not limited to, the outer cell membrane, nuclear membrane, mitochondrial and other cellular organelle membranes. Upon permeabilization with a permeabilization compound of this invention, the targeted membrane biomolecules are normally not solubilized or removed from the membrane, i.e. the membrane biomolecules remain associated with the permeabilized cell. Membrane biomolecules include, but are not limited

to, proteins, glycoproteins, lipids, carbohydrates, nucleic acids and combinations thereof, that are associated with the cellular membrane, including those exposed on the external or extracellular surface of the outer membrane as well as those that are present on the internal surface of the outer membrane, and those proteins associated with the nuclear, mitochondrial and all other intracellular organelle membranes. Membrane biomolecules also include cytoskeletal proteins.

"Genotype" or "genotyping" refers to the process of identifying intracellular genetic materials, such as DNA, that store internally coded inheritable instructions for constructing and controlling all aspects of cell life and death. "Phenotype" or "phenotyping" is defined as classifying a cell on the basis of observable outward structural elements and the production thereof (i.e. including the intermediate RNA). These include topology, morphology and other surface characteristics, all of which result from internally coded genotypic information which are incorporated into the methods of the present invention. In contrast, cell structure and integrity are not maintained during conventional RNA isolation techniques involving complete lysis of, at least, all cell structures except for nuclei and mitochondria in the presence of NP-40, usually by disintegration of all cell structures during chaotropic salt treatment and/or mechanical cellular homogenization.

Morphologic or morphology in reference to cell structure is used as customarily defined, pertaining to cell and nuclear topology and surface characteristics including intracellular or surface markers or epitopes permitting staining with histochemical reagents or interaction with detectably labeled binding partners such as antibodies. In addition morphology shall include the entire field of "morphometry" defined as: quantitative measure of chromatin distribution within the nucleus.

The terms genomic and proteomic are used as conventionally defined. "Functional" is herein used as an adjective for an empirically detectable biological characteristic or property of a cell such as "functional cellomic" which more broadly encompasses both genomic and proteomic as well as other target categories including, but not limited to, "glyconomic" for carbohydrates and "lipidomic" for cellular lipids. The resultant cell

characteristics provide profiles permitting differentiation of normal from transformed cells.

"Contacting" means bringing together, either directly or indirectly, a compound or reagent into physical proximity of a cell. The cell and/or compounds can be present in any number of buffers, salts, solutions, etc. Contacting includes, for example, placing the reagent solution into a tube, microtiter plate, microarray, cell culture flask, or the like, for containing the cell(s). The microtiter plate and microarray formats further permit multiplexed assays for simultaneously analyzing a multiplicity of cellular target compounds or components including, but not limited to, nucleic acids and proteins.

"Permeabilization compound, reagent, or composition" means any reagent that forms small pores in the cell membranes, comprising the lipid-cholesterol bilayer, while maintaining sufficient membrane, cytoplasmic and nuclear structure such that subsequent phenotypic analysis can be carried out on the permeabilized cell(s). For example, saponin is a known "pore-forming" compound that complexes with cell membrane components thereby forming numerous trans-membrane pores of about 8 nm size in the cell wall or membrane, thus allowing outward diffusion of small soluble cytosolic constituents, such as enzymes, proteins, glycoproteins, globulins, electrolytes, and the like, and internal equilibration with extracellular reagent components, such as electrolytes.

"Immunomagnetic beads" are magnetically labeled nanoparticles or microparticles also having covalently attached binding reagents (e.g. antibodies) with substantially selective affinity for surface markers or epitopes on cells, thereby achieving selective capture of magnetically labeled cells when exposed to a magnetic field such as generated in high gradient magnetic separation system (HGMS). Other terms used herein for methodologies, reagents and instruments are as conventionally defined and known to persons skilled in the art.

Preferred gene expression targets (mRNA and protein) for identifying tissue of origin, diagnosis, prognosis, therapy target characterization and monitoring include but are not limited to cells derived from cancers of the breast, prostate, lung, colon, ovary, kidney, bladder, and the like for the purpose of detection and monitoring of sensitive or resistant genes expressing